



ELSEVIER

Available online at www.sciencedirect.com

Discrete Applied Mathematics 130 (2003) 381–393

DISCRETE
APPLIED
MATHEMATICS

www.elsevier.com/locate/dam

Graphs, networks, and linear unbiased estimates

Gregory M. Constantine*

Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 30 October 2001; accepted 30 September 2002

Abstract

It is shown how a best linear unbiased estimate (blue) in the additive variety-block setting can be interpreted as a network flow, that is, a function on edges that obeys the Kirchhoff laws, of minimum square norm. An explicit expression is then obtained for the coefficients of the blue in terms of invariants of the underlying network; specifically, the invariants are: the total number of spanning trees and the number of certain selective yet specific spanning forests with just two trees. The blue is also expressed as a linear combination of bases of paths in a constructive manner. It remains a conjecture as to whether there always exists a basis of paths in which the blue is a convex combination. Consequences to design optimality are explored.

© 2003 Elsevier B.V. All rights reserved.

MSC: 05B05; 62K10; 05C20

Keywords: Path; Spanning tree; Cyclomatic number; Subspace; Estimation

1. Introduction

This work is motivated by the statistical theory of linear estimation. It specifically involves the estimation of the effects of v varieties that are compared within b blocks. In general, the blocks may be of different sizes. It is known, however, that a simple rescaling followed by an additive process allows us to assume without loss of generality that all blocks are of size two (cf. [5, p. 347], [2]). The estimation theory becomes then a problem that can be conveniently reformulated in graph theoretical language, and algebraic techniques can be used to shed light on the nature of the linear unbiased estimates that result. Statistical terminology used is nearly self-contained in

* Supported in part by a Scaife Family Foundation Grant.

E-mail address: gmc@euler.math.pitt.edu (G.M. Constantine).

what follows; further details, especially on design optimality, are found in [8]. We refer to [1] for basic results and graph theoretical notation.

In Section 2 we describe the statistical model. We then generate an abstract framework for the model by treating statistical observations as indeterminates and introduce vector subspaces to allow an algebraic formulation of the statistical estimation problem. Section 3 establishes results on the dimensionality of the subspaces of unbiased estimates related to pairs of distinct varieties. Paths between two distinct varieties are the most intuitive unbiased estimates, and we provide a constructive basis for the subspace of all (projective) unbiased estimates in terms of such paths. A study of the best linear unbiased estimates (blues) is undertaken in Section 4. The objective is to explicitly express the blue of a variety difference as a linear combination of paths in a basis. An iterative procedure for the expression of the blue is studied. Special cases, involving bases consisting of uncorrelated paths, are shown to be connected to Menger's theorem on connectivity; the blue has a particularly simple form in this case.

The concluding section offers connections to design optimality from the perspective adopted in the previous sections of the paper.

2. The additive variety-block model

There are v treatments (or varieties, as in varieties of wheat) to compare. The comparison is conducted in blocks (they may be parcels of land of possibly different fertilities). The goal is to rank the varieties (in terms of yield per acre). Each block is a subset of varieties (the type of wheat grown in it). As qualified in the Introduction, we assume without essential loss that each block contains exactly two distinct varieties. An allocation of varieties $1, 2, \dots, v$ to blocks $1, 2, \dots, b$ can be represented by an undirected graph. Vertices correspond to varieties and edges to blocks. The resulting graph may have multiple edges but no loops. Denote by y_{ij} the observation obtained on variety i in block j . We view y_{ij} abstractly as a random variable; the actual observation is merely an observed sample. The additive variety-block model postulates that the expected value of y_{ij} is equal to $\alpha_i + \beta_j$, where α_i denotes the effect of variety (or vertex) i and β_j denotes the effect of block (or edge) j . It assumes furthermore that any random variable y_{km} has the same variance σ^2 and that the variables are independently distributed. The goal is to estimate the variety effects α_i , more specifically, differences of such variety effects. The variety effects are unknown nonnegative numbers. There is generally no interest in estimating the β_j 's; they are considered nuisance parameters since their sole purpose for inclusion in the model is to be able to additively express the expected values of the y_{ij} . An easy way to dispense with them, expressed in graph theoretical language, is given below.

Consider an edge e whose end points are p and q . Form the difference $y_{pe} - y_{qe}$. The expected value of this difference is

$$E(y_{pe} - y_{qe}) = E(y_{pe}) - E(y_{qe}) = \alpha_p - \alpha_q,$$

since the β_e 's cancel out. For convenience we denote by z_e the difference $y_{pe} - y_{qe}$. Since this is slightly ambiguous, we arbitrarily orient the edges of the graph, then

unambiguously define z_e as the difference $y_{pe} - y_{qe}$, where the arrow on edge e points from q to p . Variables z_e are independently distributed.

We view the z_e 's as indeterminates. Denote by V the real vector space having the z_e 's as basis; its dimension is b , since it is equal to the number of edges in the graph. Let W be the real vector space with the basis being the unknown effects α_i , which we also view as indeterminates; its dimension is v , the number of vertices in the graph. Consider now the (formal) linear operator E from V to W , defined on the basis of V by $E(z_e) = \alpha_p - \alpha_q$, where e is an edge directed from endpoint q to endpoint p . Since we can work on each connected component of the graph, if necessary, we may assume without loss that the graph is connected. This assumption implies the existence of an undirected path P between any two distinct vertices (p and q , say) of the graph. Consider the element of V defined by $z_P = \sum \pm z_e$, where e belongs to P ; we include z_e with a plus sign in the sum if when walking from p to q without retracing, we walk on e in the direction of its arrow and use a minus sign otherwise. It is clear that $E(z_P) = \alpha_q - \alpha_p$. For a connected graph, therefore, we see that the image of the operator E has dimension $v - 1$. By the isomorphism theorem it follows that the kernel of E is a subspace of dimension $b - v + 1$ of V .

In general, we call an element x of V an (linear) unbiased estimate of w if $E(x) = w$. Denote by $\langle w \rangle$ the subspace generated by element w .

When w is an element of W , the preimage of $\langle w \rangle$ under E is a subspace of V which we call the subspace of unbiased estimates of $\langle w \rangle$. Furthermore, the subspace of unbiased estimates of $\langle \alpha_p - \alpha_q \rangle$ is denoted by V_{pq} . The variance of an unbiased estimate x of w is defined as $E[(x - w)^2]$ and is denoted by $\text{var}(x)$. Finally, a best linear unbiased estimate (blue) of w is a linear unbiased estimate of w of minimum variance. Linearity and finite dimensionality immediately insure that if w has a linear unbiased estimate, then it has a unique blue. We denote the blue of w by \hat{w} . (Since the only estimates of w we consider are elements of V , by unbiased estimate we always mean a linear unbiased estimate.) One of our aims is to understand the blue of $\alpha_p - \alpha_q$ in terms of the underlying graph structure.

3. Subspaces of unbiased estimates

Throughout this paper G is a connected directed graph (associated to a statistical experiment). We consider spaces V with basis (z_e) and W with basis (α_i) , where z_e is indexed by the edges of G and α_i by the vertices of G . By E we denote the expectation operator. We first prove the following:

Lemma 1. *The image $\text{Im } E$ is a $v - 1$ dimensional subspace of W having as basis $(\alpha_i - \alpha_j)$, for i fixed and all $j \neq i$. The kernel $\ker E$ is a $b - v + 1$ dimensional subspace of V with basis consisting of cycles.*

Proof. The only part that may need justification is the statement about $\ker E$. Any cycle (which we allow to be self-intersecting) is evidently in $\ker E$. Assume now that

$E(x) = 0$. We proceed by induction on the number of variables z_e (or edges) in x . The inductive assumption is that any vector x containing fewer than n z_e 's is a linear combination of cycles. Assume that there are n variables z_e in x . By linearity of E and the inductive assumption, it follows that x contains no cycles with fewer than n edges. If the edges in x do not form a cycle with n edges, then they form a graph without cycles. Hence there must exist a vertex j of degree 1 in x that is an end-point of edge f . But then the variable y_{jf} has expectation $\alpha_j + \beta_f$, and there is no other variable that could cancel α_j . This contradicts the fact that $E(x) = 0$. It follows that x is a scalar multiple of a cycle with n edges. Since $\ker E$ is generated by cycles we can extract a basis for $\ker E$ that consists entirely of cycles. This ends the proof. \square

To graph theorists the dimension of the kernel of E is known as the cyclomatic number. It is one of the oldest concepts in graph theory, going back to at least the work of Kirchhoff [11] on the laws of electrical current. Statisticians view this construction as a linear model, and refer to elements of $\ker E$ as error functions, since the expectation of such a function involves none of the parameters that are subject to estimation. Connections between statistical estimation in the additive setting of block designs and electrical network theory are found in the work of Tjur [12].

Our interest is in understanding the unbiased estimates of $\alpha_i - \alpha_j$, where i and j are distinct varieties. Recall that V_{ij} is the set of elements of V whose expected value is a multiple of $\alpha_i - \alpha_j$; that is, $V_{ij} = E^{-1}(\langle \alpha_i - \alpha_j \rangle)$. The next Lemma sheds further light on this issue.

Lemma 2. *The unbiased estimates V_{ij} form a subspace of V of dimension $b - v + 2$ for all $i \neq j$.*

Proof. Since the expectation of a sum of elements of V_{ij} is a scalar multiple of $\alpha_i - \alpha_j$, linearity of E insures that V_{ij} is a subspace. As to its dimension, one observes that $\ker E \leq V_{ij}$, that $V_{ij}/\ker E \cong \langle \alpha_i - \alpha_j \rangle$, and therefore the index of $\ker E$ in V_{ij} is 1. In view of Lemma 1, this ends the proof. \square

3.1. Construction of an explicit basis

In order to be as explicit as possible in our computations we construct an explicit basis for V_{ij} . The basis consists of paths between vertices i and j . Specifically, we denote the basis by B_{ij} , and is obtained as follows. Fix any spanning tree T of G . Consider any edge e that is not an edge of T . Both endpoints of $e = (p, q)$ are of degree at least 2, else the edge must be in T . Since T is a spanning tree, there is a unique path from i to p with all its edges in T , and another unique path from q to j with all its edges in T . We thus obtain a path from i to j containing edge e and with all other edges in T ; we denote it by P_e . There is also a path from i to j with all its edges in T which we label by P_0 . We now set $B_{ij} = \{P_e : e \text{ not in } T\} \cup \{P_0\}$.

Theorem 1. *The set B_{ij} is a basis for the subspace V_{ij} .*

Proof. Observe first that any path between i and j yields an unbiased estimate for $\alpha_j - \alpha_i$. Indeed, as the edges of the path are ordered, all we need to do is appropriately multiply the edges z_e 's of the path by 1 or -1 such that upon taking expectation of the weighted sum we obtain, by a telescoping effect, simply $\alpha_j - \alpha_i$. Consider the paths in B_{ij} . A path P_e with e not being an edge of T cannot be a linear combination of the other paths, since the variable z_e appears in the path P_e only. Analogously, the path P_0 cannot be a linear combination of the other paths in B_{ij} , since the existence of a nontrivial linear combination would contradict our previous sentence. This ends the proof. \square

An immediate consequence is the following:

Corollary 1. *Any unbiased estimate of $\alpha_j - \alpha_i$ is a linear combination of paths in B_{ij} with sum of coefficients equal to 1.*

Proof. As argued in the proof of Theorem 1, each path in B_{ij} is an unbiased estimate of $\alpha_j - \alpha_i$. Upon taking expectation of a linear combination of paths in B_{ij} we obtain $\alpha_j - \alpha_i$ times the sum of coefficients of the linear combination. Since this linear combination of paths is unbiased for $\alpha_j - \alpha_i$, it follows that the sum of coefficients is 1. \square

4. Best unbiased estimates

It is of interest to study the form of the blue as a linear combination of paths. The expression depends on the choice of basis B_{ij} . We thus fix an arbitrary effect $\alpha_j - \alpha_i$, the associated subspace V_{ij} , and an arbitrary basis B_{ij} of V_{ij} consisting of paths from i to j (each of which is identified with an unbiased estimate of $\alpha_j - \alpha_i$). Stack the $b - v + 2$ paths of B_{ij} in a vector $P = (P_1, \dots, P_{b-v+2})$. The problem before us now is that of finding a vector of constants C such that $E(C^t P) = (\alpha_j - \alpha_i)\mathbf{1}$, and $\text{var}(C^t P)$ is minimal. In this notation the expectation of a vector is by definition the vector of expectations of each of the individual entries; by $\mathbf{1}$ we denote the vector with all entries 1. This is a quadratic optimization problem, with the additional complexity that the entries of the vector P are not independent variables. The dependency enters through the fact that the paths share common edges, that is, common variables z_e 's. This requires the introduction of the covariance matrix of P . By definition, the covariance of random variables X and Y is the scalar $\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$.

Let F be the covariance matrix of vector P , that is, the matrix with (k, m) th entry equal to $\text{cov}(P_k, P_m)$. The covariance is related to the following graph theoretic construct. Start at vertex i on path P_k and walk toward vertex j without retracing. Each time an edge e is traced we record either z_e or $-z_e$ as to whether we walk in the direction of the arrow or against the arrow on edge e ; denote by S_k the set of these (signed) variables. For paths P_k and P_m we denote by $c^+(k, m)$ the cardinality of the set of edges that have the same variables in both paths, and by $c^-(k, m)$ the cardinality of the set of edges that have variables of opposite signs in the two paths. Observe that $c^+(k, m) + c^-(k, m) = |S_k \cap S_m|$.

Lemma 3. *The covariance $\text{cov}(P_k, P_m)$ is equal to $[c^+(k, m) - c^-(k, m)]2\sigma^2$.*

Proof. By the bilinearity of the covariance, the statement is reduced to the fact that $\text{cov}(z_e, z_e) = 2\sigma^2$ and $\text{cov}(z_e, -z_e) = -2\sigma^2$. But this directly follows from the assumptions made on the observations. This ends the proof. \square

With these facts behind us, and the notation introduced thus far, we may state as follows:

Theorem 2. *The best linear unbiased estimate of $\alpha_j - \alpha_i$ is $\hat{\alpha}_j - \hat{\alpha}_i = (1^t F^{-1} 1)^{-1} 1^t F^{-1} P$. Furthermore, $\text{var}(\hat{\alpha}_j - \hat{\alpha}_i) = (1^t F^{-1} 1)^{-1} 2\sigma^2$.*

Proof. The solution to this optimization problem is obtained in a straightforward manner by using Lagrange multipliers. Alternatively, it may also be formulated as a Gauss–Markov model by writing $E(P) = \mathbf{1}(\alpha_j - \alpha_i)$ and $\text{cov}(P) = F$; the solution then follows from the general theory of the linear model. This ends the proof. \square

Intuitively one expects that shorter paths will be weighted more than longer paths in the formula for the blue given above. This is indeed verified in the case when a basis of edge disjoint paths can be found. The number of such edge disjoint paths is found by a version of Menger’s theorem; specifically, the maximum number of edge disjoint paths between two vertices is equal to the minimum number of edges required to disconnect the two vertices. It is clear that edge disjoint paths are linearly independent. Generally, however, they are not sufficiently numerous to span a basis. When they do, we have the following:

Corollary 2. *If a basis of edge disjoint paths P_1, \dots, P_{b-v+2} of lengths $l_1 \leq \dots \leq l_{b-v+2}$ exists between vertices i and j , then the blue of $\alpha_j - \alpha_i$ is given by the convex combination*

$$\hat{\alpha}_j - \hat{\alpha}_i = \left(\sum l_k^{-1} \right)^{-1} \left(\sum l_k^{-1} P_k \right),$$

and its variance is

$$\text{var}(\hat{\alpha}_j - \hat{\alpha}_i) = \left(\sum l_k^{-1} \right)^{-1} 2\sigma^2.$$

Proof. Under these assumptions the covariance matrix F in Theorem 2 is diagonal with the l_i ’s as diagonal entries. This ends the proof. \square

Certain issues relating to the results we proved may be illustrated by an example. The example, along with Corollary 2, motivates a question regarding the convex span of paths.

Example. (a) Consider the graph G with edges $(2, 1), (2, 3), (2, 5), (3, 4), (5, 4), (5, 1)$. For simplicity of notation, to these edges we associate variables $z_2, z_5, z_3, z_6, z_4, z_1$ in the

exact corresponding order in which they are listed above. By an edge (i, j) we mean an edge with an arrow on it pointing from endpoint i to endpoint j . We intend to find the blue of $\alpha_3 - \alpha_1$. By selecting the spanning tree whose edges are all edges of G except $(5, 1)$ and $(3, 4)$, we obtain a basis of unbiased paths as follows: $P_1 = -z_1 - z_3 + z_5$, $P_2 = -z_2 + z_3 + z_4 - z_6$, $P_3 = -z_2 + z_5$. The covariance matrix of these paths is

$$F = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 4 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Following Theorem 2 we obtain that the blue, in this basis, of $\alpha_3 - \alpha_1$ is $(5P_1 + 4P_2 + 2P_3)/11$. Its variance is, again by Theorem 2, $\frac{13}{11} 2\sigma^2$. We note that in this basis the blue is a convex combination of the three paths.

(b) Let us now consider the same undirected graph G , but orient its edges differently. Specifically, the edges are $(1, 2)$, $(2, 3)$, $(2, 5)$, $(4, 3)$, $(5, 4)$, $(1, 5)$. With the same set of z variables as in (a), associated to these edges in the exact order in which they appear, we obtain another basis: $Q_1 = z_1 + z_4 + z_6$, $Q_2 = z_2 + z_3 + z_4 + z_6$, $Q_3 = z_2 + z_5$. The spanning tree we used here has all edges of G except $(1, 5)$ and $(2, 5)$. A computation yields the following inverse of the covariance matrix of the Q 's:

$$F^{-1} = \frac{1}{13} \begin{pmatrix} 7 & -4 & 2 \\ -4 & 6 & -3 \\ 2 & -3 & 8 \end{pmatrix}.$$

The blue of $\alpha_3 - \alpha_1$ in the basis of Q 's is $(5Q_1 - Q_2 + 7Q_3)/11$. As expected, it has the same variance of $\frac{13}{11} 2\sigma^2$. One notes, however, that in this latter basis the blue is no longer a convex combination of the paths.

The example illustrates several things: that the expression of the blue depends on the chosen basis of paths, that its variance is independent of the chosen basis (since it is ultimately the same linear function of the variables z_e 's, irrespective of the chosen basis), and that eventually only in some bases of paths it is a convex combination. It would be interesting to know whether a basis of paths in which the blue is a *convex combination* always exists.

Being independent of the chosen basis, it is conceivable that the variance $\text{var}(\hat{\alpha}_j - \hat{\alpha}_i) = (1^t F^{-1} 1)^{-1} 2\sigma^2$ has a purely graph theoretic interpretation. That this is indeed the case is proved in the next theorem.

Theorem 3. *The variance of the blue of $\alpha_j - \alpha_i$ is equal to $(t_{ij}/t)2\sigma^2$, where t is the number of spanning trees in G , and t_{ij} is the number of spanning trees in the graph obtained from G upon identifying vertices i and j .*

Proof. Theory of the linear model yields an expression for $\text{var}(\hat{\alpha}_j - \hat{\alpha}_i)$ in terms of the Fisher information matrix associated to the parameters of interest (cf. [10]). In

the model that we work with the Fisher information matrix is equal to $\frac{1}{2}C$, where C is the Kirchhoff, or Laplacian, matrix of the graph G . To wit, C has the degree of vertex i as the i th diagonal entry, and the negative of the number of edges between vertices i and j as the (i, j) th entry; $i \neq j$. Having row sums equal to zero, the matrix C is always singular. It has deficiency 1 in rank precisely when the graph is connected. In terms of the Fisher information matrix, the variance is expressed as $\text{var}(\hat{\alpha}_j - \hat{\alpha}_i) = (\varepsilon_j - \varepsilon_i)'(2C^-)(\varepsilon_j - \varepsilon_i)\sigma^2$, where ε_k is the column vector with all entries zero except the k th which is 1. By C^- we understand any generalized inverse of C . The expression is well defined, in the sense that it does not depend on the choice of the generalized inverse. A convenient generalized inverse is obtained as follows. Denote by C_{jj} the principal minor of the (j, j) th entry of C . The matrix C_{jj} is nonsingular of dimension $v - 1$, since G is connected. Replace the entries in the j th row and the j th column of C by zeros, and replace the entries of the minor C_{jj} by the corresponding entries of C_{jj}^{-1} , in exactly the positions in which they occur in C . Denote the resulting matrix of dimension v by C^+ ; the matrix C^+ is a generalized inverse of C . Let c^{kl} denote the (k, l) th entry of C^+ . With this choice of generalized inverse we evidently obtain

$$\text{var}(\hat{\alpha}_j - \hat{\alpha}_i) = (\varepsilon_j - \varepsilon_i)'(2C^+)(\varepsilon_j - \varepsilon_i)\sigma^2 = c^{ii}2\sigma^2. \quad (1)$$

But c^{ii} is, up to sign, equal to the cofactor of i in C_{jj} divided by the determinant of C_{jj} . By the matrix-tree theorem the determinant of C_{jj} is equal to t . On the other hand, the cofactor in question may be obtained directly from C by deleting rows i and j , and columns i and j , and evaluating the determinant of the resulting $(v - 2)$ -dimensional matrix; call this matrix M . Let G_{jj} be the graph obtained upon identifying vertices i and j of G and erasing all loops that may result. Matrix M may also be obtained by deleting the row and column of the Kirchhoff matrix of G_{ij} that is associated to the vertex in which i and j are coalesced. The matrix theorem applied to G_{ij} informs us now that the determinant of M is, up to sign, equal to the number of spanning trees t_{ij} in the graph G_{ij} . This ends the proof. \square

The expression of the blue that appears in Theorem 2 involves matrix inversion and it therefore does not offer a graph theoretical interpretation. Such an interpretation is obtained from the reduced normal equations, as is illustrated in the proof that follows. Denote by $t_{i,j,k}$ the number of forests with two trees of G with one tree containing vertices i and j and the other tree containing vertex k ; here i and j may be the same vertex but both i and j are different from k . Recall that we write $e = (i, j)$ to indicate that edge e is oriented from endpoint i to endpoint j .

Theorem 4. *The best linear unbiased estimate of $\alpha_1 - \alpha_v$ is equal to $\sum_e c_e z_e$, where the sum is over all edges of G . Explicitly, with t denoting the number of spanning trees in G , we have*

- $tc_e = (-1)^{1+i}t_{1iv} + (-1)^jt_{1jv}$ if $e = (i, j)$ with both i and j being different from v ,
- $tc_e = (-1)^{1+i}t_{1,i,v}$ if $e = (i, v)$,
- $tc_e = (-1)^it_{1,i,v}$ if $e = (v, i)$.

Proof. The reduced normal equations for the vector of parameters α is given by

$$\hat{\alpha} = \frac{1}{2} C^{-1} X_1^t (I - X_2 (X_2^t X_2)^{-1} X_2^t) Y,$$

where Y is the $2b \times 1$ vector of observations, and $X = [X_1, X_2]$ is the design matrix of the experiment. Matrix X_1 is associated to the variety effects α and is a $2b \times v$ matrix with 0 and 1 as entries, indicating the absence or presence of a variety effect in an observation. Analogously, matrix X_2 is the $2b \times b$ matrix whose (i, j) th entry is 1 if the i th entry of Y contains β_j in its expected value and is 0 otherwise. Easy calculations show that $I - X_2 (X_2^t X_2)^{-1} X_2^t = I - \frac{1}{2} (I \otimes J_2)$, where J_2 is the 2×2 matrix with all entries 1. Using these explicit matrices, it just as straightforwardly follows from the above expression of $\hat{\alpha}$ that

$$2(\hat{\alpha}_1 - \hat{\alpha}_v) = (1, 0, \dots, 0, -1) C^{-1} (d_1, \dots, d_v)^t,$$

where $d_i = \sum_e z_e$, with the sum ranging over all edges incident with vertex i . Our choice for a generalized inverse is

$$C^{-1} = \begin{pmatrix} C_{vv}^{-1} & 0 \\ 0^t & 0 \end{pmatrix},$$

with C_{vv} being the principal minor of entry (v, v) in C . Let $C_{vv}^{-1} = (c^{ij})$. With this substitution in the above equation, it follows that $2(\hat{\alpha}_1 - \hat{\alpha}_v) = \sum_{j=1}^{v-1} c^{1j} d_j$. The coefficient c^{1j} has a graph theoretical interpretation: it is equal to $(-1)^{1+j} t_{1,j,v}$, where $t_{1,j,v}$ is the number of forests of G with two trees, one of which contains vertices 1 and j and the other contains vertex v . Sorting by the edges of the graph G , we obtain the expressions written in the statement of the theorem. This ends the proof. \square

4.1. A connection to network flows

Let $\sum_e c_e z_e$ be any unbiased estimate of $\alpha_1 - \alpha_v$. The coefficients $\{c_e\}$ are a flow of value ± 1 on the oriented graph G which we now view as a network with vertex 1 as the start and vertex v as the finish (see [5, p. 191]). Indeed, the Kirchhoff law holds at every vertex, other than 1 and v , since the coefficient of α_i is 0 (under expectation) but also $\sum_e c_e$, with the sum extending over all edges that have vertex i as an end point; it follows that $\sum_e c_e z_e = 0$. Applying this observation to the best unbiased estimate, Theorem 4 yields the following:

Corollary 3. *In any graph, the forests with two trees satisfy the Kirchhoff constraints.*

It follows also from Theorem 4 that the variance of the best linear unbiased estimate of $\alpha_1 - \alpha_v$ is $2\sigma^2(\sum_e c_e^2)$, with the coefficients c_e having the specific graph theoretic interpretation mentioned in Theorem 4. Further such interpretations of the variance appear in the next section.

5. Optimality criteria

In general terms, the question of optimal statistical design asks for that planning strategy which is most informative statistically, given that a certain amount of data will become available. In the current setting, this directly translates into finding the graph which is statistically most informative among all graphs with a fixed number of vertices and a fixed number of edges. To elucidate still what we mean by statistically most informative, call the difference $\alpha_j - \alpha_i$ an elementary variety contrast. Several natural choices exist for measuring statistical information. One such choice is to take the average variance of all $\binom{v}{2}$ elementary contrasts, and seek the graph that minimizes that among all graphs with v vertices and b edges. Such a graph is called A-optimal. Another choice is the graph that yields a confidence ellipsoid of smallest Euclidean volume for a basis of elementary contrasts. A graph that does this is termed D-optimal.

Yet another choice may be the minimization of the largest variance of an elementary contrast. Generally speaking optimality criteria are functions of the eigenvalues of the Fisher information matrix. It is well-known that the D-criterion is equivalent to the maximization of the product of the nonzero eigenvalues. Analogously, the A-criterion seeks to minimize the harmonic mean of the eigenvalues. In the same spirit, a graph is E-optimal if it has the largest smallest nonzero eigenvalue, (see [9]) among all graphs with v vertices and b edges. Fiedler [7] gives interesting insights between the smallest nonzero eigenvalue of the Kirchhoff matrix and the degree of connectivity of a graph.

It is known (see [3,4]) that a D-optimal graph is one that has the maximum number of spanning trees among all graphs with v vertices and b edges. To our knowledge, no direct graph theoretical interpretation was given to the A-criterion or the MV-criterion (see [10]). The work in the previous sections allows us to formulate such an interpretation. We do, in fact, provide two separate graph theoretical formulations. The first, and perhaps most useful, is an immediate consequence of Theorem 3.

Corollary 4. *An A-optimal graph is that which minimizes $t^{-1}(\sum_{i < j} t_{ij})$ among all graphs with v vertices and b edges.*

Proof. From Theorem 3 we deduce that

$$\frac{1}{2\sigma^2} \binom{v}{2}^{-1} \sum_{i < j} \text{var}(\hat{\alpha}_i - \hat{\alpha}_j) = \binom{v}{2}^{-1} t^{-1} \left(\sum_{i < j} t_{ij} \right).$$

Since v , σ^2 , and b are fixed, this ends the proof. \square

It should be noted that Corollary 4 yields an equation for the harmonic mean of eigenvalues in terms of trees in the graph G as follows.

Corollary 5. *If $0 = \mu_0 < \mu_1 \leq \dots \leq \mu_{v-1}$ are the eigenvalues of the Kirchhoff matrix, then*

$$(v-2) \sum_{k=1}^{v-1} \frac{1}{\mu_k} = t^{-1} \left(\sum_{i < j} t_{ij} \right).$$

Proof. From Corollary 4, using Eq. (1), we obtain

$$\begin{aligned} t^{-1} \left(\sum_{i < j} t_{ij} \right) &= \frac{1}{2\sigma^2} \sum_{i < j} \text{var}(\hat{x}_i - \hat{x}_j) = \sum_{i < j} (a^{ii} + a^{jj} - 2a^{ij}) \\ &= (v-2)\text{Trace}(C^-) = (v-2) \sum_{k=1}^{v-1} \frac{1}{\mu_k}, \end{aligned}$$

where $C^- = (a^{ij})$ is the Moore–Penrose generalized inverse of C . We use the fact that both C and C^- have zero row sums. This ends the proof. \square

We make a connection now between the average variance criterion and graph polynomials. A tree cover of G is a partitioning of the vertices such that each class of the partition is a tree. If to a tree with t vertices we attach the indeterminate w_t and to a cover the product of the indeterminates of the trees it contains, then the polynomial

$$T(G; \mathbf{w}) = \sum_c \prod_{t \in c} w_t \quad (2)$$

is called the tree polynomial of G ; the sum is over all tree covers c of G . We refer to [6] for a general theory of graph polynomials.

The polynomial T written in (2) above verifies a fundamental recurrence

$$T(G; \mathbf{w}) = T(G'; \mathbf{w}) + T(G''; \mathbf{w}), \quad (3)$$

where G' is the graph obtained from G upon deletion of an edge (call the edge ab), and G'' is the graph obtained from G by identifying vertices a and b and omitting any loops formed. It is easy to see that Eq. (3) is true by partitioning the tree covers into those that contain edge ab and those that do not. If we count the number of its vertices plus the number of its edges as size of a graph, then both G' and G'' are graphs of size smaller than G . Formula (3) allows us therefore to inductively compute the tree polynomial of any graph. Conversely, and more relevantly, starting out with a graph H and adding an edge ab between two of its vertices (corresponding to a comparison of two varieties within an edge), we obtain the graph $G = H \cup \{ab\}$. Graphs G and H have the same number of vertices, G having an additional edge. By (3) we have $T(G; \mathbf{w}) = T(H \cup \{ab\}; \mathbf{w}) = T(H; \mathbf{w}) + T(\bar{H}; \mathbf{w})$, where \bar{H} is the graph obtained from H by identifying vertices a and b and deleting the resulting loops. This shows how the tree polynomial can be built sequentially by the addition of edges.

The relevance of the tree polynomial to statistical efficiency is seen by relating it to the Kirchhoff polynomial. By the Kirchhoff polynomial $K(G; x)$ of G we understand the characteristic polynomial of the matrix Kirchhoff matrix C . It is helpful to realize that the coefficient of x^{v-i} in $K(G; x)$ is equal to the i th symmetric sum of eigenvalues of C (denoted by S_i). The average variance is (up to the multiple $v-2$, which we may ignore since v is fixed) equal to the harmonic mean of eigenvalues of C , as is stated in Corollary 3. In terms of the symmetric sums the harmonic mean equals S_{v-2}/S_{v-1} . The average variance is thus measured by the quotient of the coefficient of x^2 to that of x

in $K(G; x)$. On the other hand, as we shall see next, K is obtained by a substitution in T , and therefore the coefficients in question can be interpreted as tree covers.

Indeed, if we replace w_t by tx in $T(G; \mathbf{w})$ what results is $K(G; x)$. This can be seen as follows. The coefficient of x^i in $K(G; x)$ is S_{v-i} . Upon substituting, the coefficient of x^i becomes $\sum_c \prod_{j=1}^i n_j$, where the sum is over all covers c with i trees of sizes n_1, \dots, n_i ; $n_j \geq 1$, $\sum jn_j = v$. To see that the two coefficients of x^i are in fact the same, add a new vertex $v+1$ to G and connect it to all the other vertices of G ; denote the resulting graph \tilde{G} . Label each of the edges originating from $v+1$ by the indeterminate x . Up to the cardinalities of the trees in the covers, there is a bijection between the tree covers of G and the spanning trees of \tilde{G} : the trees of the cover being associated to the endpoints of the edges originating at the vertex $v+1$. In particular, the covers with i trees of G correspond to spanning trees of \tilde{G} that contain i edges labelled x . By the matrix-tree theorem the number of spanning trees in \tilde{G} is the determinant of $C - xI$, the principal minor that results upon deleting the row and column corresponding to vertex $v+1$. The determinant of $C - xI$ is simply the characteristic polynomial of C ; thus the coefficient of x^i is S_{v-i} . On the other hand, this coefficient counts the number of spanning trees in \tilde{G} with exactly i edges labelled x which (as we mentioned) equals the number of spanning forests with i trees of G weighted by the cardinalities of the trees, i.e., it equals $\sum_c \prod_{j=1}^i n_j$. This shows that $S_{v-i} = \sum_c \prod_{j=1}^i n_j$ and that $K(G; x)$ is obtained upon substituting tx for w_t in the tree polynomial $T(G; \mathbf{w})$. We summarize what was discussed as follows:

Theorem 5. *The average variance of the $\binom{v}{2}$ elementary contrasts is equal to $[(v-2)(\sum n_j(v-n_j))/vt]2\sigma^2$, where t is the number of spanning trees in the graph G , and the sum extends over all covers of G with exactly two trees, one with n_j and the other with $v-n_j$ vertices.*

While the usual interpretation of the average variance is in terms of the spectrum of the Fisher information matrix, the interpretation given in Theorems 3 and 4 is in terms of the graph itself. This presents the advantage of having greater control over efficiency, since, in general, the eigenvalues are intractable functions of the entries of the information matrix. Recurrence (3), via the substitution $w_t = tx$, provides the means by which such direct control can be exercised.

References

- [1] N.L. Biggs, Algebraic Graph Theory, Cambridge University Press, Cambridge, 1974.
- [2] C.S. Cheng, Optimality of certain asymmetrical experimental designs, Ann. Statist. 6 (1978) 1239–1261.
- [3] C.S. Cheng, Maximizing the total number of spanning trees in a graph: two related problems in graph theory and optimum design theory, J. Combin. Theory Ser. B 31 (1981) 240–248.
- [4] G.M. Constantine, On the optimality of block designs, Ann. Inst. Statist. Math. 38 (1986) 161–174.
- [5] G.M. Constantine, Combinatorial Theory and Statistical Design, Wiley, New York, 1987.

- [6] E.J. Farrell, On a general class of graph polynomials, *J. Combin. Theory Ser. B* 26 (1979) 111–121.
- [7] M. Fiedler, Algebraic connectivity of graphs, *Czechoslovak Math. J.* 23 (1973) 298–305.
- [8] A.S. Hedayat, D. Majumdar, A-optimal block designs for control-test treatment comparisons, *Technometrics* 26 (1984) 363–370.
- [9] M. Jacroux, On the E-optimality of regular graph designs, *J. Roy. Statist. Soc. B* 42 (1980) 205–209.
- [10] M. Jacroux, On the determination and construction of MV-optimal block designs for comparing test treatments with a standard treatment, *J. Statist. Plann. Inference* 15 (1987) 205–225.
- [11] G. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, *Ann. Phys. Chem.* 72 (1847) 497–508 (English translation: *Trans. Inst. Radio Engrs. CT* 5 (1958) 4–7).
- [12] T. Tjur, Block designs and electrical networks, *Ann. Statist.* 19 (2) (1991) 1010–1027.